DOCUMENT RESUME

ED 142 851 CG 011 552

AUTHOR Samuel, William; And Others
TITLE Test Environment, Sex of Tester, and the IQ of Female Subjects.
PUB DATE [76]
NOTE 28p.; Paper presented at the Annual Meeting of the American Psychological Association (84th, Washington, D.C., September 3-7, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Intelligence Tests; Junior High School Students; *Motivation; *Performance; *Racial Factors; Research Projects; *Sex Differences; Socioeconomic Status; *Testing

ABSTRACT

The performance subscales to the Wechsler Intelligence Scale for Children were administered to 208 black and 208 white female junior high school students. In addition to race and socioeconomic status (above or below the median) of subject, the variables of race and sex of tester, test atmosphere (evaluative or gamelike), and tester expectation (high or low) were placed in a 2x2x2x2x2x2 factorial design. Atmosphere, expectation, and socioeconomic status interacted significantly with the race of the subject in determining mean IQ. With the exception of some results gathered by white testers that deviated in an intriguing manner from the predicted pattern, the findings generally replicated those reported for male subjects in an earlier paper. Performance was presumed to be mediated by differing levels of testee motivation, such that achievement was greatest in conditions which were supposed to generate a moderate level of motivational arousal. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED142851

Test Environment, Sex of Tester, and the IQ of Female Subjects

William Samuel, Sonja Grayson, Beverly Sherman,
Debra H. Soto, and Katye Winstead

California State University, Sacramento

ABSTRACT

The performance subscales to the Wechsler Intelligence Scale for Children were administered to 208 black and 208 white female junior high school students. In addition to race and socioeconomic status (above or below the median) of subject, the variables of race and sex of tester, test atmosphere (evaluative or gamelike), and tester expectation (high or low) were placed in a 2x2x2x2x2x2 factorial design. Atmosphere, expectation, and socioeconomic status interacted significantly with the race of the subject in determining mean IQ. With the exception of some results gathered by white testers that deviated in an intriguing manner from the predicted pattern, the findings generally replicated those reported for male subjects in an earlier paper. Performance was presumed to be mediated by differing levels of testee motivation, such that achievement was greatest in conditions which were supposed to generate a moderate level of motivational arousal.

CG 011552

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William Samuel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

Test Environment, Sex of Tester,

and the IQ of Female Subjects

Why do individuals and socioeconomic groupings of individuals sometimes differ in their performance on cognitive tasks? One customary set of answers to this question has emphasized the types of skills or deficits inculcated by different sorts of family and cultural experiences (see, e.g., Garcia, 1972; Scarr and Weinberg, 1976; Zajonc, 1976), while another tradition has concentrated on fairly durable qualities of individual temperament and mental proficiency (e.g., Jensen, 1969; Eysenck, 1973; Loehlin, Lindzey, and Spuhler, 1975). The social psychology of the test situation itself may, however, exert an influence on individual and group performance which is to some extent independent of the foregoing factors (Sattler, 1970). In a recent article, Samuel, Soto, Parks, Ngissah, and Jones (1976) have in fact reported that a student's IQ score may be altered by the setting in which the performance subscales to the Wechsler Intelligence Scale for Children (WISC) are administered.

It was assumed by the latter investigators that differences in mean IQ between treatment conditions were mediated by variations in internal arousal induced by the experimental manipulations. Specifically, they noted that several investigators have found internal arousal to bear an inverted U-shaped relationship with performance (Duffy, 1957; Hokanson and Burgess, 1964; Doerr and Hokanson, 1965; Doob and Kirshenbaum, 1973). That is, arousal facilitates performance up to some optimal, moderate level but thereafter becomes disruptively distracting and, hence, performance-debilitating. Thus, subjects habitually low in test anxiety or motivational arousal may actually show higher achievement when placed in a stressful rather than a

relaxed setting (Sarason, 1961; McCoy, 1965); chronically over-aroused subjects, on the other hand, are the ones whose performance is most likely to be impaired by a stressful examination environment (Sarason, 1961; Weiner and Samuel, 1975).

To permit a test of this U-shaped arousal hypothesis, Samuel, et al. administered the WISC to 416 junior high and high school students equally divided by race and sex. Black or white male experimenters described the subscales as either evaluative tests of intelligence or creative games and playthings; they also forecast either an above- or below-average performance for each subject based on a pretest score and various "school records." Regardless of the race of the tester or the race of the subject, it was anticipated that in the evaluative atmosphere the ego-threatening and competitive nature of the setting would induce a state of fairly high anxiety in testees and that if the experimenter then expressed a high expectation this might reduce anxiety to a more moderate level and so facilitate performance. If the tester instead expressed a <u>low</u> expectation in the evaluative atmosphere, the added stress of his criticism should definitely debilitate performance. In the gamelike atmosphere, by contrast, the setting was intended to be so relaxed that testee motivation would be insufficient for peak performance, particularly if the reassurance of a high expectation was provided. Here it was assumed that achievement would be facilitated rather than debilitated by the moderate anxiety induced by the tester's low expectation. In short, peak performance in response to praise was predicted for the evaluative atmosphere (an instructional effect) while peak performance in response to criticism was predicted for the gamelike (a <u>counter</u>instructional effect). An Atmosphere x Expectation interaction

in the IQ data supported these predictions; subject self-ratings of mood tended to confirm the predicted pattern of motivational arousal.

There were two major exceptions to the Atmosphere x Expectation interaction discussed above. Black males and white females in the presence of a white tester in the evaluative atmosphere showed counterinstructional effects rather than the predicted instructional effect. The black males were, of course, in the presence of a same-sex but opposite-race tester, while for white females the tester was same-race but opposite-sex. We then recruited an additional sample of 209 male subjects, which yielded a male population of sufficient size to permit socioeconomic status (SES) to be treated as an independent variable. When this was done, it was found that it was primarily black males of low SES who were responsible for the counterinstructional effect in the evaluative atmosphere with a white tester and that low SES white males also showed these tendencies, though to a lesser degree. Ignoring the effects of the atmosphere and expectation manipulations, it was found that all groups of subjects, regardless of race, sex, or SES, scored higher in IQ in the presence of a white tester.

At this point, it was decided to replicate the experiment using an additional sample of 208 female subjects but to employ female rather than male testers. The results for 416 female subjects tested by male or female experimenters are reported in the present paper. The variables of test atmosphere, tester expectation, race of tester, and race of subject were manipulated, as before, and sex of tester and SES of subject (above or below the median) became the last two factors in a 2x2x2x2x2x2 design.

It was, of course, anticipated that the Atmosphere x Expectation interaction would replicate for female subjects. We might also have predicted,

on the basis of our earlier results and data gathered by other investigators (e.g., Katz, 1970), that mean IQs would generally be higher for groups tested by white experimenters. The data on this topic are highly contradictory, however; sometimes whites as well as blacks perform better in the presence of a black examiner (Kennedy and Vega, 1965; McClelland, 1974), while other times subjects of each racial group do their best in the presence of a same-race tester (Allen, Dubanoski, and Stevenson, 1966; Katz, Henchy, and Allen, 1968). Consequently, we did not make any a priori predictions for the Race of Tester variable. It was also difficult to specify in advance what effect the tester's sex might have on the achievement of female testees. While some investigators (e.g., Rosenthal, 1966, pp. 42-56) report that testees perform better in the presence of an opposite-sex rather than a same-sex examiner, others have found that, for female subjects at least, a same-sex examiner facilitates performance (Pedersen, Shinedling, and Johnson, 1968; Shinedling and Pedersen, 1970; Montanelli and Hill, 1969).

METHOD

The performance subscales to the WISC were administered to 416 female junior high and high school students, equally divided by race; all were under 16 years of age. Arriving at an office provided by the school, the subject encountered Experimenter$_1$ of a team of two experimenters. Two of the teams had all-male experimenters; on one of these, both experimenters were white while on the other both were black. The two remaining teams were all-female, one consisting of white and the other of black experimenters. The experimenters were graduate and advanced undergraduate students in psychology.

Experimenter$_1$ described the subscales as either evaluative tests of intelligence or as creative games and playthings. To enhance the effectiveness of this atmosphere manipulation, male testers wore a jacket and tie in the evaluative condition but removed the jacket and loosened the tie when the atmosphere was to be gamelike. The female testers wore fashionable suits in the evaluative atmosphere but removed the suit jacket and draped it casually over a chair in the gamelike setting.

Experimenter$_1$ also forecast either an above- or below-average performance for each subject on the basis of an alleged consultation of various school records. To reinforce this expectation manipulation, the subject was administered the Object Assembly subtest of the WISC, after which it was announced that she had either done rather well, above the average, or rather poorly, below the average. In the evaluative setting, an imposing interval timer was running during the period the subject assembled each picture.[1] In the gamelike atmosphere, the examiner glanced at a wall clock and expressed regret at having to time the activities due to the necessity for keeping things on schedule. When the subtest was completed, Experimenter$_1$ explained that an associate had a few other tests (or games) for the student to work with. In departing, Experimenter$_1$ put on his or her jacket if the atmosphere had been gamelike or placed the timer on the floor if the setting had been evaluative. After a brief pause, Experimenter$_2$, who was blind as to the subject's prior treatment, arrived and administered the four remaining WISC performance subscales. Experimenter$_2$'s verbal interaction with the subjects was limited to the very brief instructions required for each subtest; these were extracted word-for-word from the WISC manual (Wechsler, 1949).

Following completion of the subscales, subjects rated the quality of their performance on a 7-unit scale running from "very poorly" to "very well." Next, each completed a mood adjective checklist (Nowlis, 1965). The adjectives on the checklist comprise scales for measuring aggression (e.g., "angry"), anxiety ("clutched up"), concentration ("engaged in thought"), egotism ("boastful"), elation ("overjoyed"), fatigue ("tired"), sadness ("sorry"), skepticism ("suspicious"), and surgency ("playful"). For each adjective, the subject indicated the degree to which it described her feelings on a 4-unit scale running from "definitely not" to "definitely." Finally, subjects provided their home address and parents' occupations. The addresses were referred to census tracts and the occupations to tables of ranked occupational categories (Reiss, Duncan, Hatt, and North, 1961, pp. 109-138) to yield a 100 point index of socioeconomic status (SES).

Especial care was taken to see that each participant left the research situation in a pleasant frame of mind. A probe for suspicion was followed by a reassuring debriefing and payment of a $3 remuneration. The subject was then sworn to secrecy and released. Further details regarding this and other aspects of the procedure may be found in Samuel, et al. (1976).

RESULTS AND DISCUSSION

Success of the Experimental Manipulations

The atmosphere and expectation manipulations appear to have been successful. Self-ratings of performance were considerably more positive for subjects in the high than in the low expectation conditions ($F(1,412) = 317.52$, $p < .001$). On the mood checklist, subjects rated themselves significantly more elated in the high than in the low expectation conditions ($F(1,412) = 26.94$, $p < .01$); they also described themselves as more elated

in the gamelike than in the evaluative atmosphere ($F(1,412) = 4.36$, $p < .05$). Self-ratings of anxiety were higher in the evaluative than in the gamelike atmosphere, but this difference fell short of significance ($F(1,412) = 1.73$).

IQ Data

The IQ scores reported below are derived from the four subtests administered by $Experimenter_2$, prorated according to procedures in the WISC scoring manual (Wechsler, 1949). The prorating caused calculated IQs to be slighly higher than they would have been if the discarded Object Assembly score had instead been included and no prorating applied.

Table 1 shows the mean and standard deviation of the IQs measured in each of the 64 cells of the experimental design. Since cell ns varied as a result of the division of subjects into groups high and low in SES, an unweighted-means analysis was applied to the data (Winer, 1962). The

Insert Table 1 about here

overall mean and standard deviation for IQ were 106.03 and 12.54, respectively, values which conform reasonably well with those obtained in previous research (Jensen, 1969). The mean IQ for whites was 111.98 while that for blacks was 100.08, an interracial difference of 11.90 points ($F(1,352) = 64.75$, $p < .001$). Regardless of the race of the testees, the IQ of students of high SES ($M = 109.65$) was reliably above that of their low SES counterparts ($M = 102.40$; $F(1,352) = 9.13$, $p < .01$).

While the race of the tester did not have a reliable overall impact on IQ ($F < 1$), a main effect for Sex of Tester indicated that subjects performed somewhat better in the presence of a female ($M = 109.33$)

rather than a male experimenter ($\underline{M}$ = 102.73; $\underline{F}(1,352) = 14.83$, $p < .01$). A significant Race of Tester x Sex of Tester interaction revealed, however, that the latter result was due primarily to exceptionally high scores in the presence of a black female tester ($\underline{M}$ = 111.63; $\underline{F}(1,352) = 12.45$, $p < .01$). The most parsimonious interpretation of these results would appear to be that the black female testers were consistently overly generous in timing the performance of testees and that the sex of the tester *per se* did not have an overall effect on performance.

The main effects for race and SES of subject are not at all surprising given the persistence with which they have appeared in the testing literature. What was of greatest importance to the present investigation was the predicted Atmosphere x Expectation interaction, which would have indicated a motivational influence on performance, regardless of the race or SES of the examinee. As can be seen in Table 2, however, this interaction failed to reach

Insert Table 2 about here

significance ($.10 < p < .20$). Clearly, though, the pattern of mean IQs was the same as that previously reported for male subjects (the latter findings are shown in the far right-hand column of Table 2). In the data for males, the predicted instructional effect in the evaluative atmosphere and the counterinstructional effect in the gamelike were both reliable ($ps < .05$), which was interprted as indicating that subjects were under-aroused in the totally reassuring high expectation--gamelike condition and over-aroused in the highly stressful low expectation--evaluative setting.

A reliable Atmosphere x Expectation x SES x Race of Subject interaction was observed in the data for female subjects and is shown in Figure 1

($F(1,352) = 4.30$, $p < .05$). The generally U-shaped appearance of these

Insert Figure 1 about here

curves--which developed from the predicted tendency toward instructional effects in the evaluative atmosphere and counterinstructional effects in the gamelike--suggests that the Atmosphere x Expectation interaction might have been reliable but for the behavior of low SES blacks, who showed a non-significant counterinstructional effect in the evaluative atmosphere. Interestingly, in the previously-published data for male subjects, it was also low SES blacks who showed a counterinstructional effect in the evaluative atmosphere, particularly when in the presence of a white tester. Still, the effect was not in that case powerful enough to dilute the significance of the Atmosphere x Expectation interaction.

In an effort at discovering why low SES blacks showed a counter-instructional effect in the evaluative atmosphere, a procedure was followed which had been found to be useful in interpreting the data for male subjects: namely, analyzing separately the IQ scores gathered by black and white testers.[3] As can be seen in Table 1, when the tester was black, deviations from the pattern of means characteristic of the Atmosphere x Expectation interaction were observed for high SES whites in the presence of a male tester and for high SES blacks and low SES whites in the presence of a female tester. Some of this variability was due to the fact that cell ns differed considerably (ranging from 2 to 11) when subjects tested by male or female examiners were split at the overall median for SES.[4]

When cell ns were made more uniform by combining the data gathered by black male and black female testers, the results for female subjects

were strikingly similar to those previously reported for male subjects tested by a black examiner, as can be seen in Figure 2. The generally

Insert Figure 2 about here

U-shaped appearance of these curves reflects the fact that the Atmosphere x Expectation interaction was reliable for female ($p < .05$; see Table 2) as well as male ($p < .05$) subjects. The curve for low SES blacks showed only a very slight tendency toward a counterinstructional effect in the evaluative atmosphere.

In the data for subjects tested by white male and white female examiners, lack of agreement with the pattern of means characteristic of the Atmosphere x Expectation interaction was the rule rather than the exception (see Table 1). The major deviations from this predicted pattern consisted in counterinstructional effects in the evaluative atmosphere; these were primarily manifested by high SES whites and, to a lesser extent, high SES blacks with a male tester and by low SES blacks and, to a lesser extent, low SES whites with a female tester.

Reducing the variability in cell n by combining the results gathered by white male and female examiners yielded U-shaped curves for only two groups of subjects: low SES whites and high SES blacks. When these results for female subjects were compared with the previously-published data for male subjects tested by white examiners in Figure 3, the curves

Insert Figure 3 about here

for low SES whites and high SES blacks appeared quite similar. The curves for high SES white and particularly for low SES black females deviated markedly from the U-shaped pattern as well as from the results

for their male counterparts.

A reliable ($p < .05$) Atmosphere x Expectation x SES x Sex of Tester interaction in the data gathered by white testers reflected the foregoing observations and is shown in Table 3. High SES subjects with a male tester

Insert Table 3 about here

showed a very marginal counterinstructional effect in the evaluative atmosphere with a male tester ($t = 1.30$, $p < .20$); their counterparts in the presence of a female tester, by contrast, showed the predicted instructional effect ($t = 1.87$, $p < .06$).[5] Low SES subjects in the presence of a white tester showed a significant counterinstructional effect in the evaluative atmosphere with a female tester ($t = 2.66$, $p < .05$) and a non-significant instructional effect when the tester was male ($t < 1$). The race of the subject did not reliably interact with these results, but it should be noted nonetheless that the latter counterinstructional effect was highly significant for blacks ($t = 3.09$, $p < .002$) but not for whites ($t < 1$). Thus, to return to the question which was posed by the data in Figure 1, it appears that the counterinstructional effect in the evaluative atmosphere among low SES blacks was displayed mainly by subjects in the presence of a white female tester.

One final question to be asked is whether it is possible to identify any motivational mediators of the interaction shown in Table 3. Self-rated aggression was the only likely possiblity suggested in exploratory analyses of the mood checklist data. While the Atmosphere x Expectation x SES x Sex of Tester interaction was not reliable for aggression in the presence of a white tester ($F(1,192) = 1.25$), the slightly less complex Expectation x SES x Sex of Tester interaction was of marginal significance

($F(1,192) = 2.87$, $p < .10$). The latter interaction was also observed in the analysis of the IQ data shown in Table 3 ($F(1,176) = 4.85$, $p < .05$). In general, a high degree of self-rated aggression in the presence of a white tester was associated with relatively poor performance on the WISC ($r = -.20$; $t = 2.93$, $p < .01$); this relationship was somewhat stronger in the evaluative ($r = -.26$; $t = 2.69$, $p < .02$) than in the gamelike setting ($r = -.15$, n.s.).

Particularly in the evaluative atmosphere, the pattern of mean aggression scores was quite intriguing. For high SES subjects with a white female experimenter and low SES subjects with a white male experimenter, anger was, logically enough, greater when the tester expressed a low ($M = 2.68$ and 1.92, respectively) rather than a high expectation ($M = 1.71$ and 1.70, respectively), and WISC performance was debilitated as a consequence. For high SES subjects with a white male experimenter or low SES subjects with a white female experimenter, on the other hand, aggression was greatest when the tester expressed a high ($M = 3.22$ and 2.60, respectively) rather than a low expectation ($M = 1.85$ and 0.86, respectively), and it was in the former treatments that performance was worst. Why would high SES females have become irritated by a white male tester's high expectation while low SES subjects were put off by a white female tester's high expectation? The data provide no definitive answer to this question. Perhaps the female subjects run in these settings felt they were for some reason being patronized by the tester and that the latter's encouraging expectations were somehow insincere.

Conclusion

It must be acknowledged that the present research replicated the previously-observed superiority in mean IQ of white vs. black and of high vs. low SES testees (Jensen, 1969; Duncan, Featherman, and Duncan, 1972). It

is clear in Figure 1, however, that these main effects for race and social class of subject reliably interacted with the variables of test atmosphere and tester expectation in influencing mean IQ. Furthermore, it is entirely conceivable that stress reactions to testing situations and expectations for performance can develop into durable personality traits over the course of an individual's lifetime. If we had precise measures of such chronic reactions and expectancies we might well have been surprised by the powerful control they exerted over performance and the strength of their interactions with race and social class. As it was, we placed some acute manipulations of motivational arousal in an experimental design and found that they had at least a limited impact on observed IQ.

The predicted Atmosphere x Expectation interaction was marginally supported in the overall analysis of variance. This interaction was clearly significant when the tester was black ($p < .05$), and the data for female subjects in the presence of a black tester closely resembled previously-published data for males run under these conditions (see Figure 2). The Atmosphere x Expectation interaction received only directional support when the tester was white, mainly because of the occurrence of counter-instructional effects in the evaluative atmosphere among high SES subjects in the presence of a male tester and low SES subjects in the presence of a female tester. These complicating factors produced a reliable Atmosphere x Expectation x SES x Sex of Tester interaction in the data gathered by white examiners (see Table 3).

Surprisingly, the race of the tester had no overall impact on performance, and while there was a main effect indicating that subjects did better in the presence of a female rather than a male tester ($p < .01$),

this result seems to have been an artifact of exceptionally high achievement when the tester was a black female. Therefore, the safest conclusion would appear to be that neither the race nor the sex of the tester per se had a simple, main effect influence on performance. There are indications in Table 2, of course, that the race of the tester affected the strength of the Atmosphere x Expectation interaction, and Table 3 suggests that the sex of the tester influenced the occurrence of counter-instructional effects in the evaluative atmosphere among subjects tested by a white experimenter.

While some uncertainty remains as to the exact reasons for the failure of the Atmosphere x Expectation interaction to attain significance when the tester was white, it seems reasonable to infer from the present data that the atmosphere and expectation manipulations did have a reliable, though limited, influence on WISC performance among all racial and socio-economic groups in the sample. The significant variations in mean IQ which were produced by these manipulations may for the most part be interpreted as manifestations of a systematic alteration of the examinee's state of motivational arousal between experimental conditions.

REFERENCES

Allen, S. A., Dubanoski, R. A., and Stevenson, H. W. Children's performance as a function of race of E, race of S, and type of verbal reinforcement. Journal of Experimental Child Psychology, 1966, 4, 248-256.

Doob, A. N., and Kirshenbaum, H. M. The effects on arousal of frustration and aggressive films. Journal of Experimental Social Psychology, 1973, 9, 57-65.

Doerr, H. O., and Hokanson, J. E. A relation between heart rate and performance in children. Journal of Personality and Social Psychology, 1965, 2, 70-76.

Duffy, E. The psychological significance of the concept of 'arousal' or 'activation'. Psychological Review, 1957, 64, 265-275.

Duncan, O. D., Featherman, D. L., and Duncan, B. Socioeconomic background and achievement. New York: Seminar Press, 1972.

Eysenck, H. J. (Ed.) The measurement of intelligence. Baltimore, Md.: Williams and Wilkens, 1973.

Garcia, J. IQ: The conspiracy. Psychology Today, 1972, 6 (9), 40ff.

Hays, W. L. Statistics for psychologists. New York: Holt, 1963.

Hokanson, J. E., and Burgess, M. Effects of physiological arousal level, frustration, and task complexity on performance. Journal of Abnormal and Social Psychology, 1964, 68, 698-702.

Jensen, A. R. How much can we boost IQ and scholastic achievement? Harvard Educational Review, 1969, 39, 1-123.

Katz, I. Experimental studies of Negro-white relationships. In L. Berkowitz (Ed.) Advances in experimental social psychology. Vol. 5. New York: Academic Press, 1970.

Katz, I., Henchy, T., and Allen, H. Effects of race of tester, approval-disapproval, and need on Negro children's learning. Journal of Personality and Social Psychology, 1968, 2, 839-843.

Kennedy, W. A., and Vega, M. Negro children's performance on a discrimination task as a function of examiner race and verbal incentive. Journal of Personality and Social Psychology, 1968, 8, 38-42.

Loehlin, J., Lindzey, G., and Spuhler, J. N. Race differences in intelligence. San Francisco, Calif.: Freeman, 1975.

McClelland, L. Effects of interviewer-respondent race interactions on household interview measures of motivation and intelligence. Journal of Personality and Social Psychology, 1974, 29, 392-397.

McCoy, N. Effects of test anxiety on children's performance as a function of instructions and type of task. Journal of Personality and Social Psychology, 1965, 2, 634-641.

Montanelli, D. S., and Hill, K. T. Children's achievement expectations and performance as a function of two consecutive reinforcement experiences, sex of subject, and sex of experimenter. Journal of Personality and Social Psychology, 1969, 13, 115-128.

Nowlis, V. Research with the mood adjective checklist. In S. S. Tomkins and C. E. Izard (Eds.), Affect, cognition, and personality. New York: Springer, 1965.

Pedersen, D. M., Shinedling, M. M., and Johnson, D. L. Effects of sex of examiner and subject on children's quantitative test performance. Journal of Personality and Social Psychology, 1968, 10, 251-254.

Reiss, A. J., Duncan, O. D., Hatt, T. K., and North, C. C. Occupations and social status. Glencoe, Ill.: Free Press, 1961.

Rosenthal, R. Experimenter effects in behavioral research. New York: Holt, 1966.

Samuel, W., Soto, D. J., Parks, M., Ngissah, P., and Jones, B. Motivation, race, social class, and IQ. Journal of Educational Psychology, 1976, 68, 273-285.

Sarason, I. G. The effect of anxiety and threat on the solution of a difficult task. Journal of Abnormal and Social Psychology, 1961, 62 165-167.

Sattler, J. Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73, 137-160.

Scarr, S., and Weinberg, R. A. IQ test performance of black children adopted by white families. American Psychologist, 1976, 31, 726-739.

Shinedling, M. M., and Pedersen, D. M. Effects of sex of tester and student on children's gain in quantitative and verbal performance. Journal of Psychology, 1970, 76, 79-84.

Wechsler, D. WISC manual: Wechsler intelligence scale for children. New York: The Psychological Corporation, 1949.

Weiner, M., and Samuel, W. The effect of attributing internal arousal to an external source upon test anxiety and performance. Journal of Social Psychology, 1975, 96, 255-265.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Zajonc, R. B. Family configuration and intelligence. Science, 1976, 192, 227-236.

FOOTNOTES

1. The timer was a prop employed to enhance the effectiveness of the atmosphere manipulation; in practice, the experimenters used a wall clock to time the subjects' performance in all conditions.

2. In all of the figures, facilitation of visual scanning is the purpose of the lines used to connect the data points common to a given racial, sexual, or socioeconomic grouping of subjects; it is in no way being suggested that the categories on the abscissa constitute a continuous variable.

3. Two marginally reliable interactions involving the race of the tester lent some statistical support to this procedure: Expectation x SES x Race of Tester x Sex of Tester ($F(1,352) = 2.73$, $p < .10$) and Atmosphere x Expectation x SES x Race of Subject x Race of Tester x Sex of Tester ($F(1,352) = 3.42$, $p < .07$).

4. Contributing to this variability in cell n was the fact that 74% of the 208 subjects tested by male experimenters were run during 1972-73 and 26% during 1973-74; all of the subjects tested by female experimenters were run during 1973-74. Samuel, et al. (1976) noted that the two schools from which subjects were recruited during 1972-73 were located in less prosperous neighborhoods than were those utilized in 1973-74. Consequently, students tested by male experimenters tended to be lower in SES than were those tested by female experimenters. Just 36% of the former were above the overall median in SES whereas 64% of the latter were above the median. Because the split was at the overall median, however, SES was equivalent within high or low SES groups, respectively, regardless of the sex of the tester.

5. These and all other t tests were performed according to procedures for planned comparisons in Hays (1963, pp. 464-471). The harmonic mean was used to estimate n for all tests.

Table 1

Mean IQ for Female Students

SES	Atmosphere	Expectation		Male Tester: Black Subjects: Black Tester	Male Tester: Black Subjects: White Tester	Male Tester: White Subjects: Black Tester	Male Tester: White Subjects: White Tester	Female Tester: Black Subjects: Black Tester	Female Tester: Black Subjects: White Tester	Female Tester: White Subjects: Black Tester	Female Tester: White Subjects: White Tester
High	Evaluative	High	$\bar{x}$	103.00	99.75	109.14	110.20	106.75	115.00	123.20	114.45
			s.d.	18.03	9.22	12.23	15.47	12.14	9.90	10.02	14.53
			n	3	4	7	5	8	10	10	11
		Low	$\bar{x}$	88.00	101.67	109.50	122.00	111.17	100.11	114.30	108.50
			s.d.	17.86	17.61	13.48	11.46	12.64	14.41	8.73	7.93
			n	5	6	4	7	6	9	10	10
	Gamelike	High	$\bar{x}$	82.50	89.00	116.00	117.00	101.83	97.62	119.00	113.89
			s.d.	3.53	9.49	13.65	10.66	13.53	13.04	6.44	11.81
			n	2	4	6	6	6	8	5	9
		Low	$\bar{x}$	110.00	93.00	114.67	118.33	110.67	101.71	119.90	112.87
			s.d.	9.85	14.07	14.09	7.57	15.91	10.92	16.35	17.54
			n	3	4	6	3	6	7	10	8
Low	Evaluative	High	$\bar{x}$	94.20	96.33	104.00	111.50	104.80	93.33	113.33	106.00
			s.d.	14.79	19.62	14.59	16.40	13.05	1.16	6.51	16.97
			n	10	9	6	8	5	3	3	2
		Low	$\bar{x}$	94.00	89.43	99.33	112.33	103.00	117.25	114.67	107.00
			s.d.	17.99	6.63	21.28	15.16	17.42	9.36	15.70	11.27
			n	8	7	9	6	7	4	3	3
	Gamelike	High	$\bar{x}$	90.55	106.44	92.14	108.29	104.86	95.40	114.50	109.25
			s.d.	10.25	9.07	12.82	5.94	11.89	8.96	18.58	10.56
			n	11	9	7	7	7	5	8	4
		Low	$\bar{x}$	94.00	99.44	103.43	115.20	105.00	96.67	109.67	108.40
			s.d.	18.32	10.74	14.00	17.01	15.25	10.61	7.51	9.37
			n	10	9	7	10	7	6	3	5

Table 2

Mean IQs Contributing to the

Atmosphere x Expectation Interaction

Atmosphere	Expectation	Female Subjects: Black Tester[a]	Female Subjects: White Tester[b]	Female Subjects: Combined[c]	Male Subjects (Data from Samuel, et al., 1976)[d]
Evaluative	High	107.52	107.88	107.70	108.50
	Low	103.83	106.54	105.18	104.16
Gamelike	High	103.04	105.65	104.35	103.91
	Low	107.87	105.90	106.88	107.75

[a]There were 52 subjects per cell; $F_{\text{Atmos x Expect}}(1,176) = 4.13$, $p < .05$.

[b]There were 52 subjects per cell; $F_{\text{Atmos x Expect}}(1,176) < 1$.

[c]There were 104 subjects per cell; $F_{\text{Atmos x Expect}}(1,352) = 2.10$, $.10 < p < .20$.

[d]There were 104 subjects per cell; $F_{\text{Atmos x Expect}}(1,384) = 6.50$, $p < .02$.

Table 3

The Atmosphere x Expectation x SES x Sex of Tester Interaction for Mean IQ in the Presence of a White Tester[a]

SES	Atmosphere	Expectation		Male Tester	Female Tester
High	Evaluative	High	$\bar{x}$	105.56	114.71
			n	9	21
		Low	$\bar{x}$	112.62	104.53
			n	13	19
	Gamelike	High	$\bar{x}$	105.80	106.24
			n	10	17
		Low	$\bar{x}$	103.86	107.40
			n	7	15
Low	Evaluative	High	$\bar{x}$	103.47	98.40
			n	17	5
		Low	$\bar{x}$	100.00	112.86
			n	13	7
	Gamelike	High	$\bar{x}$	107.25	101.56
			n	16	9
		Low	$\bar{x}$	107.74	102.00
			n	19	11

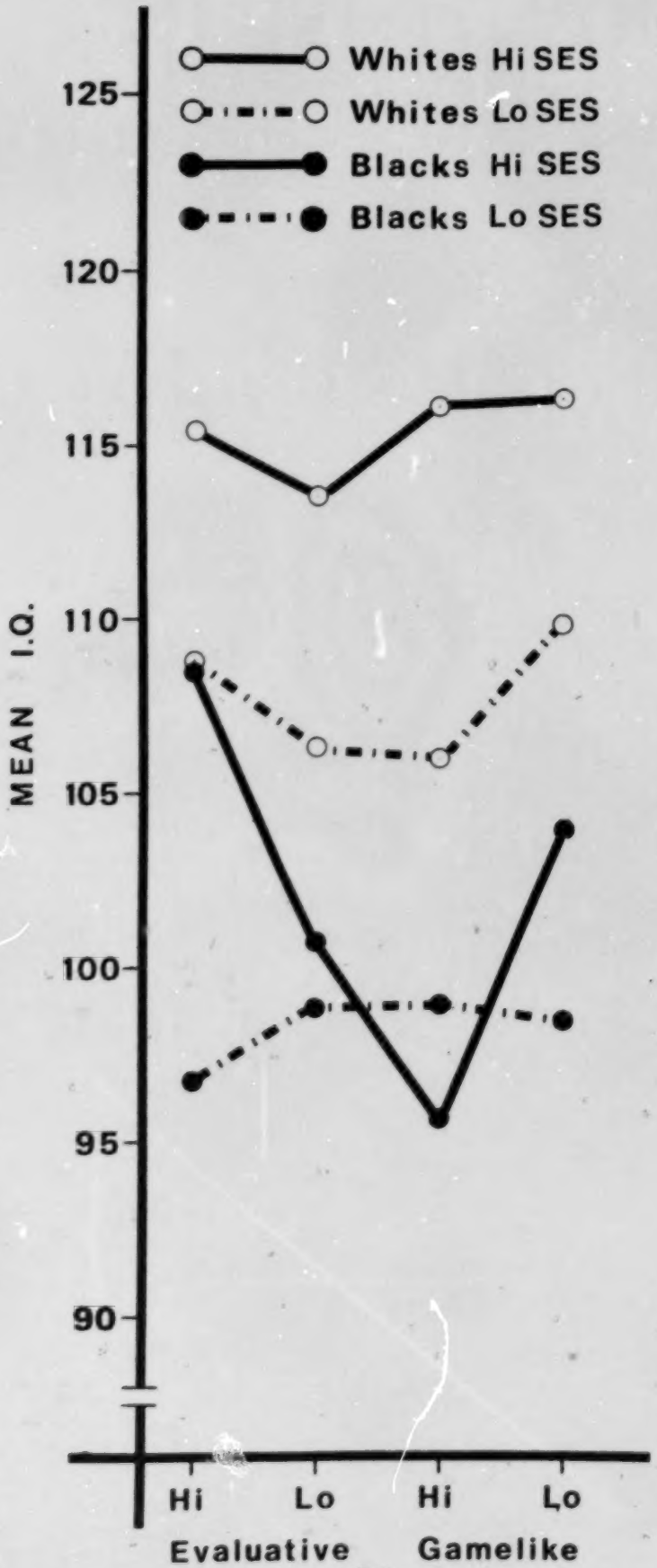
[a] $F(1,176) = 4.10$, $p < .05$.

FIGURE CAPTIONS

Figure 1. Mean IQs contributing to the Atmosphere x Expectation x SES x Race of Subject interaction.

Figure 2. Mean IQ for female or male subjects in the presence of black testers (the data for males are from Samuel, et al., 1976).

Figure 3. Mean IQ for female or male subjects in the presence of white testers (the data for males are from Samuel, et al., 1976).


Whites Hi SES
Whites Lo SES
Blacks Hi SES
Blacks Lo SES
125
120
115
110
105
100
95
90
MEAN I.Q.
Hi
Lo
Hi
Lo
Evaluative
Gamelike

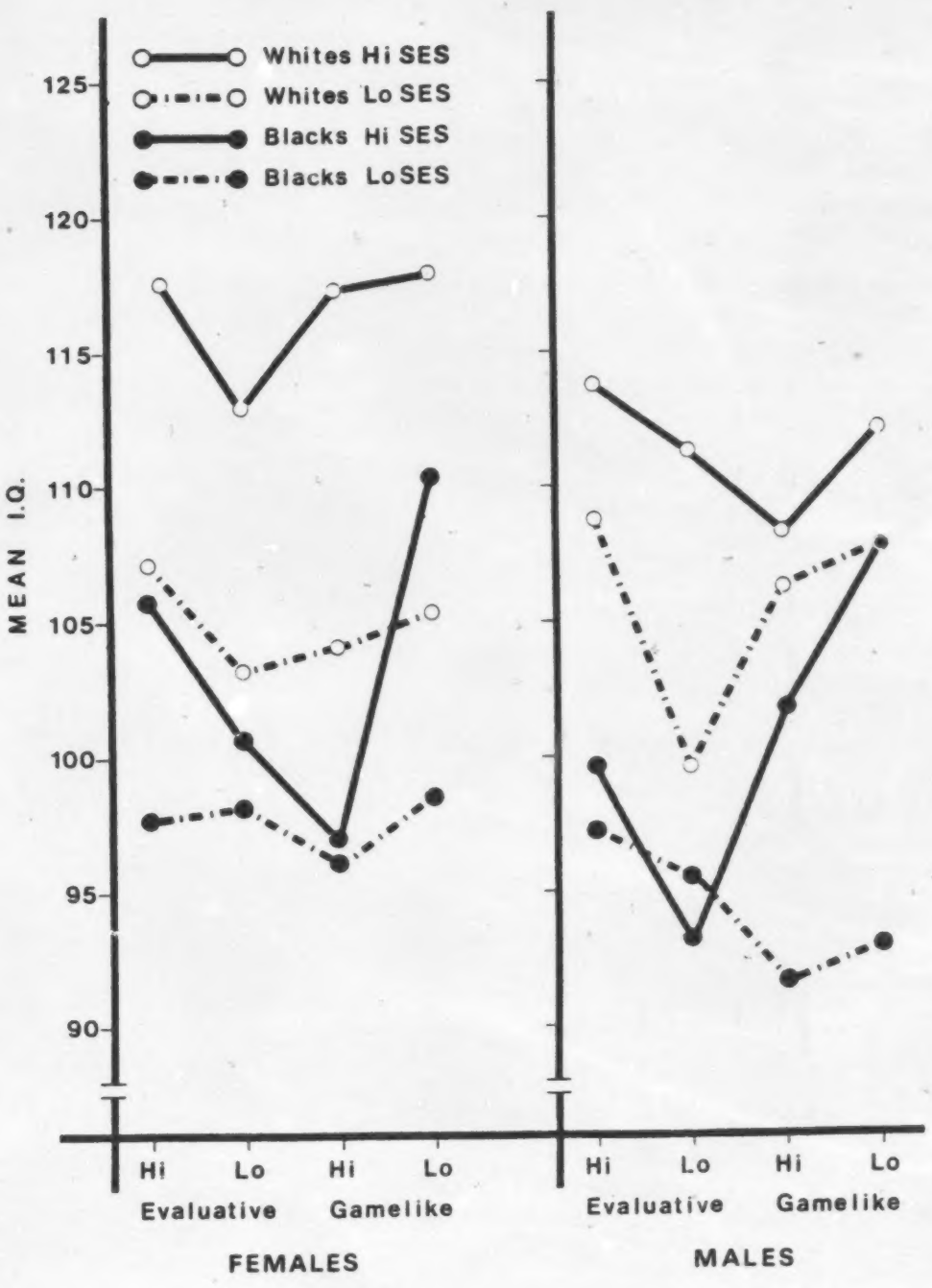

Whites Hi SES
Whites Lo SES
Blacks Hi SES
Blacks Lo SES
MEAN I.Q.
125
120
115
110
105
100
95
90
Hi
Lo
Hi
Lo
Evaluative
Gamelike
FEMALES
Hi
Lo
Hi
Lo
Evaluative
Gamelike
MALES

Figure 2

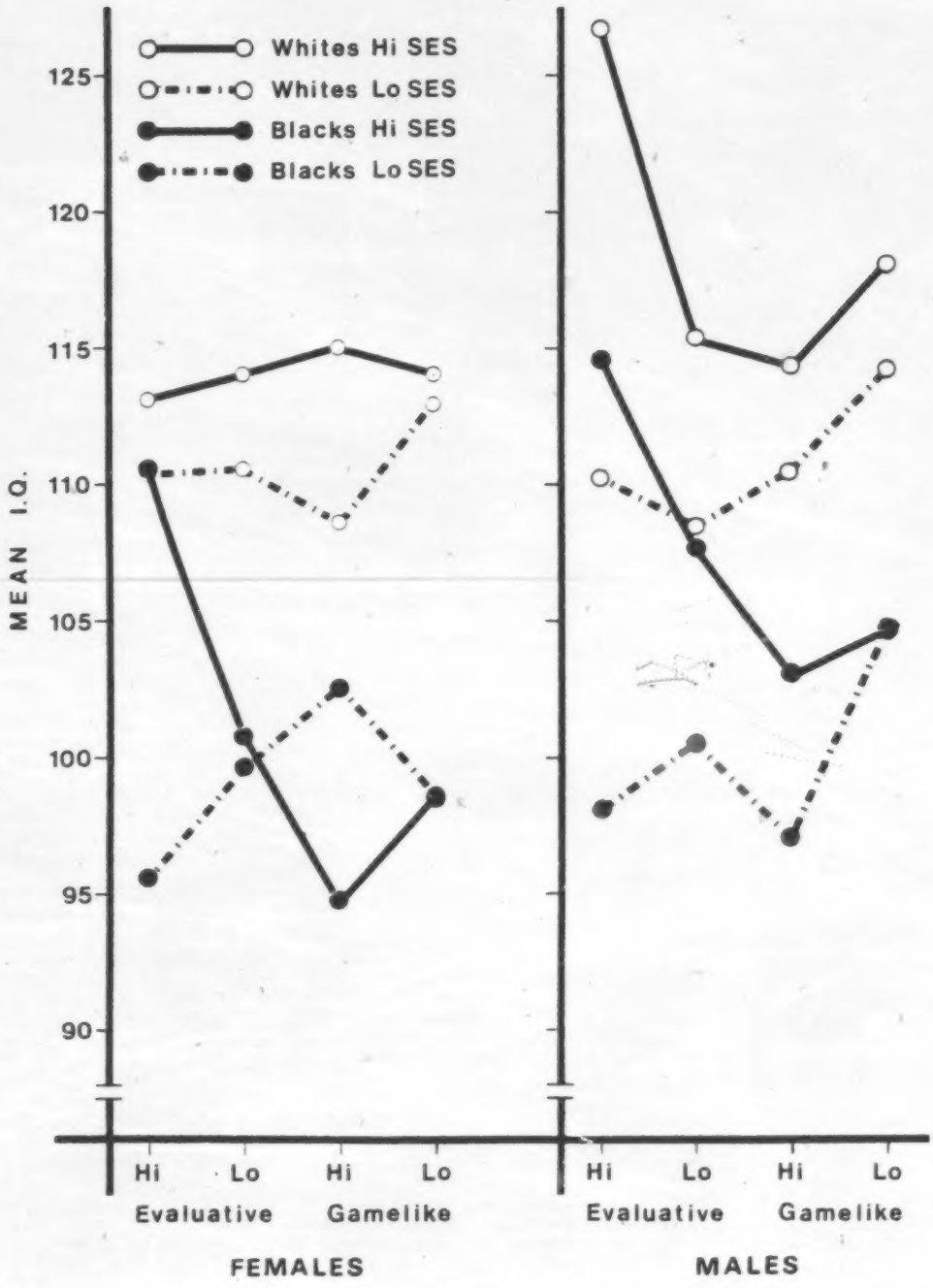

Whites Hi SES
Whites Lo SES
Blacks Hi SES
Blacks Lo SES
MEAN I.Q.
125
120
115
110
105
100
95
90
Hi
Lo
Hi
Lo
Evaluative
Gamelike
FEMALES
Hi
Lo
Hi
Lo
Evaluative
Gamelike
MALES

Figure 3